

ORAL PRESENTATION

Open Access

High-throughput sequencing of the DBA/2J mouse genome

Xusheng Wang¹, Richa Agarwala², John A Capra³, Zugen Chen⁴, Deanna M Church², Daniel C Ciobanu⁵, Zhengsheng Li¹, Lu Lu¹, Khyobeni Mozhui¹, Megan K Mulligan¹, Stanley F Nelson⁴, Katherine S Pollard³, Williams L Taylor¹, Donald B Thomason¹, Robert W Williams^{1*}

From UT-ORNL-KBRIN Bioinformatics Summit 2010
Cadiz, KY, USA. 19-21 March 2010

Background

The DBA/2J mouse is not only the oldest inbred strain, but also one of the most widely used strains. DBA/2J exhibits many unique anatomical, physiological, and behavior traits. In addition, DBA/2J is one parent of the large BXD family of recombinant inbred strains [1]. The genome of the other parent of this BXD family—C57BL/6J—has been sequenced and serves as the mouse reference genome [2]. We sequenced the genome of DBA/2J using SOLiD and Illumina high throughput short read protocols to generate a comprehensive set of ~5 million sequence variants segregating in the BXD family that ultimately cause developmental, anatomical, functional and behavioral differences among these 80+ strains.

Results

We generated approximately 13.2 and 38.9× whole-genome short reads of DBA/2J females using Illumina GA2 and ABI SOLiD massively parallel DNA sequencing platforms. Comparing to the C57BL/6J reference genome sequence, we identified over 4.5 million single nucleotide polymorphisms (SNPs), including 84 non-sense and ~11,000 missense mutations, 78% of which are novel. We also detected ~568,000 insertions and deletions (indels) within single short reads and ~9,400 between mate-paired reads. Approximately 300 inversions were detected by SOLiD mate-pair reads, 46 of which span at least one exon. In addition, we identified ~22,000 copy number variants (CNVs) in the range of 1 Kb to 100 Kb (Figure 1).

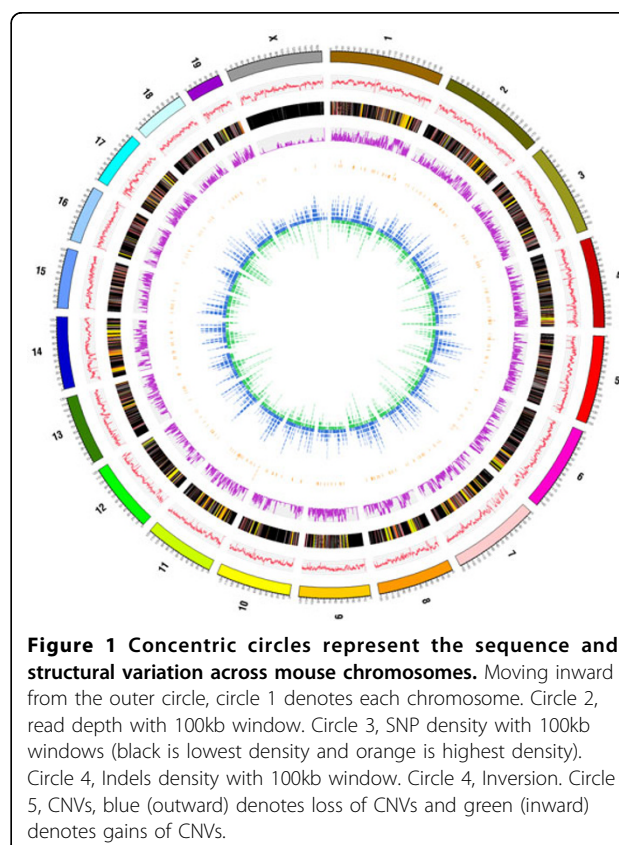


Figure 1 Concentric circles represent the sequence and structural variation across mouse chromosomes. Moving inward from the outer circle, circle 1 denotes each chromosome. Circle 2, read depth with 100kb window. Circle 3, SNP density with 100kb windows (black is lowest density and orange is highest density). Circle 4, Indels density with 100kb window. Circle 5, CNVs, blue (outward) denotes loss of CNVs and green (inward) denotes gains of CNVs.

Conclusion

Our study generates the first consensus sequence for the DBA/2J and creates a compendium of sequence and structural variations that will be used by the community of researchers who study complex traits in mouse models. The sequence data provide a novel resource with which to initiate reverse genetic analysis of complex

* Correspondence: rwilliam@nb.uthsc.edu

¹University of Tennessee Health Science Center, Memphis, TN 38163, USA

traits, particularly by exploiting strong alleles (premature stop codons, frame-shift mutations, and deletion) that differentially affect members of the BXD strain family. The DBA/2J genome is also an essential prerequisite to unbiased alignment of RNA-seq and ChIP-seq data generated using BXD strains and any other cross involving these two common parental strains.

Author details

¹University of Tennessee Health Science Center, Memphis, TN 38163, USA. ²National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA. ³Gladstone Institutes, University of California, San Francisco, CA 94158, USA. ⁴University of California, Los Angeles, CA 90095, USA. ⁵University of Nebraska, Lincoln, NE 68588, USA.

Published: 23 July 2010

References

1. Peirce JL, Lu L, Gu J, Silver LM, Williams RW: **A new set of BXD recombinant inbred lines from advanced intercross populations in mice.** *BMC genetics* 2004, **5**:7.
2. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P: **Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420**(6915):520-562.

doi:10.1186/1471-2105-11-S4-O7

Cite this article as: Wang *et al.*: High-throughput sequencing of the DBA/2J mouse genome. *BMC Bioinformatics* 2010 **11**(Suppl 4):O7.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

